Title

Method of Monitoring the Number of Amide Bonds in Peptides Using Near-Infrared Spectroscopy

Author(s)

Mika Ishigaki, Atsushi Ito, Risa Hara, Shun-ichi Miyazaki, Kodai Murayama, Keisuke Yoshikiyo, Tatsuyuki Yamamoto, and Yukihiro Ozaki

Journal

Anal. Chem. 2021, 93, 5, 2758–2766

Published

December 23, 2020

URL

https://pubs.acs.org/doi/10.1021/acs.analchem.0c03424

A method of monitoring the number of amide bonds in peptides using near-infrared spectroscopy

Mika Ishigaki[1,2]*, Atsushi Ito[3], Risa Hara[3], Shun-ichi Miyazaki[3]*, Koudai Murayama[3], Keisuke Yoshikiyo[1], Tatsuyuki Yamamoto[1,2], Yukihiro Ozaki[4]

[1]*Institute of Agricultural and Life Sciences, Academic Assembly, Shimane University, 1060 Nishikawatsu, Matsue, Shimane, 690-8504, Japan*

[2]*Raman Project Center for Medical and Biological Applications, Shimane University, 1060 Nishikawatsu, Matsue, Shimane 690-8504, Japan*

[3]*Research and Development Department, Yokogawa Electric Corporation, 2-9-32 Nakacho, Musashino, Tokyo 180-8750, Japan*

[4]*School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan*

*Authors to whom correspondence should be sent.

*E-mail: ishigaki@life.shimane-u.ac.jp (M.I.)

shun-ichi.miyazaki@yokogawa.com (S.M)

Abstract

Using near-infrared (NIR) spectroscopy, we aimed to develop a method of monitoring the increasing number of amide bonds with the elongation of the chain length of peptides. Since peptide synthesis can be monitored by evaluating the increasing number of amide bonds with dehydration occurring between amino acids, polyglycine, which has the simplest structure among polyamino acids, was studied, and the key bands whose absorption intensities increased with the elongation of the chain length, such as the bands attributed to glycine, diglycine, triglycine, and tetraglycine, were searched. The bands due to the combinations of the amide A and amide II/III modes in the region of 5000-4500 cm^{-1} were revealed to be good candidates for key bands; their second derivative intensities increased as the number of amide bonds increased, regardless of pH, solvent species, and the presence of protecting groups. The number of amide bonds was evaluated by a partial least square regression (PLSR) using the abovementioned combination bands, and a calibration model with a high determination coefficient ($\geq$0.99) was constructed. These results not only have demonstrated the usefulness of NIR spectroscopy as a process analytical technology (PAT) tool for the

process of synthesizing peptide in a micro flow reactor but also have provided basic knowledge for analyzing amide bonds in the NIR spectra of proteins, polyamino acids, polypeptides, and polyamides.

Introduction

Pharmaceuticals with a sufficient permeability through cell membranes, a high selectivity for the target, a high binding ability, and no side effects are ideal. Two major categories of pharmaceuticals exist: small molecule-based drugs and antibody drugs. While they, of course, have pharmaceutical advantages, there are still some remained disadvantages that should be solved. Special peptide drugs, on the other hand, have recently attracted attention as a next generation breakthrough pharmaceutical, since they are expected to possess all the abovementioned advantageous properties.[1]

A microflow reactor is a promising candidate for the synthesis of special peptide drugs.[2-7] Chemical products are generally made using a conventional method of chemical synthesis named a “batch method". In this method, all raw materials are put into a reaction kettle, and the objective products are taken out after the reactions are completed. By repeating the operation,

products with complicated structures can be synthesized through multiple steps. However, the isolation and purification of synthetic intermediates at each step are needed, and much waste is inevitably generated.[3, 7-9] In addition, chemical reactions performed in large kettles have problems, such as the heterogeneity of heat transfer and stirring efficiency. In the method using a microflow reactor, on the other hand, target products can be obtained by flowing the raw material stepwise through the column without the isolation and purification of intermediates (Figure 1).[2, 3, 6, 7] Furthermore, since the homogeneity of the reaction system can be ensured, and the reaction time can be controlled in milliseconds by controlling the flow path length and flow rate, and the stable synthesis process can be constructed.[2, 5-7] The United States Food and Drug Administration proposed process analytical technology (PAT) in the last two decades,[10, 11] and they recommended the replacement of the batch method with the continuous flow method.[6, 7] To establish an industrial production line with the method with a microflow reactor, technology used for monitoring chemical reactions over time and in multiple ways is essential.

Spectroscopic techniques have recently received keen interest as

powerful tools of PAT,[11-17] and near-infrared (NIR) spectroscopy is one of the most compelling techniques among them. NIR spectroscopy is mainly concerned with the absorption and reflection of light in the NIR region (12500 – 4000 cm^{-1}).[18-21] NIR light has a high permeability because its energy corresponds to the absorption energy of the overtones and combinations of fundamentals that are forbidden transitions in terms of the harmonic oscillation approximation. Therefore, NIR spectroscopy can be used to measure spectra of targets in a nondestructive and noninvasive manner, as this method does not require preprocessing. Furthermore, since NIR light can be transmitted through an optical fiber, on-line and in-line remote sensing can be constructed in manufacturing processes.[11] Thus, NIR spectroscopy has been applied as a tool of pharmaceutical monitoring for evaluating the manufacturing process and final products and has also been applied to other various fields, such as bioscience, agriculture, food engineering, and polymer engineering.[13-16, 18-23]

In the NIR spectra of proteins and peptides, several bands due to overtones and combinations of an amide group are observed.[18-24] The amide group has mainly three fundamental vibrational modes, amide I, II, and III.

In the NIR region, bands due to the second overtone of amide I and combination modes of N-H stretching and amide II or amide III modes can be observed.[23-27] Since the bands due to amide groups shift with the secondary structural changes of proteins, the secondary structural changes of proteins, such as from an α–helix to a β–sheet, can be studied using NIR spectroscopy.[25, 26, 28-30] However, analyzing NIR spectra can be difficult because of their complexity due to overlapping bands caused by overtones and combinations of fundamentals. Advanced analytical methods of spectra, such as chemometrics, are often required.[18, 19] Furthermore, absorbance peaks due to amide groups to some extent depend on, for example, protein species. An investigation of amide bands using NIR spectroscopy is not straightforward, and the current knowledge of amide band analysis is not sufficient to apply the method to the practical PAT manufacturing of peptide drugs. Therefore, we aimed to develop a monitoring method for peptide synthesis using NIR spectroscopy by exploring variations in the NIR absorption bands with the elongation of the chain length of compounds such as glycine, diglycine, triglycine, and tetraglycine.

In the present study, we searched for key absorption bands that change

with the increase in the number of amide bonds formed by dehydration condensation between amino acids and estimated the quantitative accuracy of the number of amide bonds using the key bands. It was found that the key bands whose intensities varied with varying the elongation of the chain length were those due to the combination of amide A and III modes and those of amide A and II modes, regardless of the different conditions (i.e., regardless of whether the reaction proceeded in aqueous solutions with different pH values, in an organic solvent, or with or without protection groups). Furthermore, diglycines with an additional amino acid (GlyGlyX) which has a characteristic structure, such as an additional amide bond or an aromatic ring in its side chain, were also examined for the quantitative analysis of the number of amide bonds. The results revealed that the number of amide bonds can always be evaluated with a determination coefficient of approximately 0.99 using the bands arising from the combinations of amide A and amide II/III modes.

Even though the reaction of amide bond formation is one of the most important reactions in organic chemistry, amide bond formation can be sometimes overlooked because of its very widespread use. If the behavior of

NIR absorption bands due to amide bonds can be fully understood, the basic knowledge has an explosive potential for applications in protein chemistry, food chemistry, biochemistry, polymer industry, and so on, and its ripple effect is extremely large. The present results not only have demonstrated the usefulness of NIR spectroscopy as a PAT tool for studying the process of peptide synthesis in a micro flow reactor but also have provided new basic knowledge for analyzing amide bands in the NIR spectra of proteins, polyamino acids, polypeptides, and polyamides.

Materials and Methods

A glycine stock solution was prepared by dissolving 100 g of glycine (MW=75.07, Sigma Aldrich Co.) in ultrapure water to achieve a volume of 500 mL. Different volumes of aqueous HCl (35-37.0% w/w, FUJUFILM Wako Pure Chemical Co., Japan) and NaOH (50 w/v%, FUJUFILM Wako Pure Chemical Co., Japan) were added to the 10 mL stock solution to change the pH, and 18 samples were filled with ultrapure water to achieve a final volume of 20 mL. The pH of the stock solution was 6.31, and the pH range of the other 17 samples was 0.20-13.43. A series of polyglycines including

glycine (G1), diglycine (G2), triglycine (G3), and tetraglycine (G4) were dissolved in 0.5 mol/L NaOH (199-02185, FUJUFILM Wako Pure Chemical Co., Japan), 0.5 mol/L HCl (086-01105, FUJUFILM Wako Pure Chemical Co., Japan), and ultrapure water to achieve a concentration of 200 mM.

Organic solutions of 9-fluorenylmethyloxycarbonyl group (Fmoc)-protected G(n)-OH (n=1-3), Fmoc-protected G(n) with 6-polyethylene glycol monomers (6-PEG) (n=1-5), N-tert-butoxycarbonyl (Boc)-protected G(n)-OH (n=1-3), and Boc-protected diglycine with X (X=Ala, Ser(tBu), Glu(OtBu), Lys(Boc), Phe, and Tyr(tBu)) were prepared using dimethyl sulfoxide (DMSO, Guaranteed Reagent, FUJUFILM Wako Pure Chemical Co., Japan) at concentrations of 50, 100, 150 and 200 mM. Detailed information about the polyglycine chemical reagents is presented in Supporting Information (SI) 1.

For the NIR measurements, a Spectrum One FT-NIR system (Perkin-Elmer, Waltham, MA, USA) was used. The optical path length of the quartz cell used was 1 mm. The wavenumber region measured was 10000-4000 cm^{-1}, the spectral resolution was 8 cm^{-1}, and 256 spectra were accumulated for the aqueous samples. For the DMSO solutions, the spectral resolution and

the number of spectra accumulated were set at 4 cm^{-1} and 128, respectively. The temperature of the quartz cell was controlled by a bath circulator (NESLAB RTE7, Thermo Scientific, Waltham, MA, USA) operated at 25 and 30 °C for the aqueous and organic solution samples, respectively. The stability of the temperature controller was ±0.1 °C. For the spectral analysis, the chemometrics software Unscrambler X 10.2 (Camo Analytics, Oslo, Norway) and OriginPro 6.1 (OriginLab Corporation, Massachusetts, USA) were used.

Results and discussion

Aqueous solutions of polyglycine

Figures 2A and 2B depict NIR absorbance spectra obtained in the 5000-4200 cm^{-1} region for glycine aqueous solutions with a pH in the range of 0.20-13.43 and their second derivatives, respectively. Two peaks were observed at approximately 4440 and 4320 cm^{-1} in the absorbance spectra (Figure 2A). In the second derivative spectra, three peaks except for the abovementioned two bands were located at 4960, 4730, and 4550 cm^{-1} (Figure 2B). The two peaks at 4440 and 4320 cm^{-1} were common to all the pH values, even though

there was a pH-dependent peak shift, and these peaks are assigned to the combination of the C-H stretching and bending modes.[24, 31] The three peaks at 4960, 4730 and 4550 cm^{-1}, on the other hand, appeared only in the spectra of alkaline solutions, and their second derivative intensities increased with increasing pH values. Amino acids contain both an amino and a carboxyl group, showing two dissociation equilibrium steps, as shown in Equations (1) and (2).

$$NH_3^+-CH_2-COOH \rightleftharpoons NH_3^+-CH_2-COO^- + H^+ \quad (1)$$

$$NH_3^+-CH_2-COO^- \rightleftharpoons NH_2-CH_2-COO^- + H^+ \quad (2)$$

The dissociation constants (K_{a1}, K_{a2}) are defined in Equation (3) and (4), and the pK_a values are given as pK_{a1}=2.34 and pK_{a2}=9.60 at 25 ℃.[32]

$$K_{a1} = \frac{[NH_3^+CH_2COO^-][H^+]}{[NH_3^+CH_2COOH]} \quad (3)$$

$$K_{a2} = \frac{[NH_2CH_2COO^-][H^+]}{[NH_3^+CH_2COO^-]} \quad (4)$$

Since the functional group existing only in the pH range above 9.60 was the NH_2 group, the three peaks at 4960, 4730 and 4550 cm^{-1} were derived from the NH_2 group. This result indicated that drastic spectral changes occur upon going from NH_3^+ to NH_2.

The second derivative spectra of a series of polyglycines (G1, G2, G3, and G4) in 0.5 mol/L NaOH solutions are depicted in Figure 3A. As the number of glycines increases by one, the numbers of both the amide bond and methylene group increase by one. Under the alkaline conditions, three peaks at 4960, 4730 and 4550 cm^{-1}, which were due to the NH_2 group, were commonly detected. The second derivative intensities of the peaks at 4380 and 4248 cm^{-1} increased with the elongation of the chain length, that is, with the increase in the number of amide bonds. Since these two bands at 4380 and 4248 cm^{-1} were assigned to the combinations of the C-H stretching and bending modes of the methylene group of glycine, the result that the second derivative intensities increased with the elongation of the chain length was reasonable. Furthermore, a new spectral variation also appeared at approximately 4618 cm^{-1}. The peak at 4618 cm^{-1} is due to the combination of amide A and amide III modes.[23, 25] The intensity variation at 4618 cm^{-1} likely corresponds to the number of amide bond changing to 1, 2, and 3 in G2, G3, and G4, respectively.

The second derivative of the peak at 4618 cm^{-1} seemed to line up at an equal interval, and thus, its peak intensity could be a good candidate to

evaluate the number of amide bonds. The second derivative spectra of the same series in 0.5 mol/L HCl solutions and pure water are shown in Figure 3B and 3C, respectively. Since G4 with a concentration of 200 mM could not be dissolved in pure water, the corresponding spectrum of G4 is not shown in Figure 3C. The spectral variations in the 4450-4200 cm^{-1} region that arose from the combination of C-H stretching and bending modes seemed to depend on the polyglycine chain length. In the present study, the origin of the spectral variation is not discussed in detail, but the coordinate bond of the hydrogen ion may affect the vibration of the methyl group on the *N*-terminus site, and the effect diminished with the elongation of the main polyglycine chain.

The change in the intensity of the peak at 4618 cm^{-1} followed the same trend as the number of amide bonds in both the acidic solutions and pure water, as in the case of the alkaline solution (Figure 3). The intensity of the peak at 4890 cm^{-1}, which depends on the chain length, on the other hand, changed, as could be seen in the acidic solutions and pure water but not in the alkaline solutions. This band is due to the combination of amide A and amide II modes.[23, 24, 30] In the case of alkaline solutions, the strong band

attributed to the NH_2 group appeared at 4960 cm^{-1}, and its tail seemed to disturb the trend of the amide contribution. Therefore, it became clear that the number of amide bonds could be quantitatively estimated using the combination of the amide A and amide III modes at approximately 4618 cm^{-1} for the spectra obtained in the aqueous solutions.

DMSO solutions of Fmoc- and Boc-protected polyglycine

To find a similar key band for the quantitative assessment of the number of amide bonds in organic solvents, a series of Fmoc-protected polyglycine (Fmoc-G(n)-OH, n=1, 2, 3) and Boc-protected polyglycine (Boc-G(n)-OH, n=1, 2, 3) in DMSO was also investigated. The chemical structures of Fmoc-G2-OH and Boc-G2-OH are depicted in Figures 4A and 4B as examples, respectively. Figures 5A and 5B depict the second derivative spectra of the NIR spectra in the 5000-4500 cm^{-1} region for Fmoc-protected and Boc-protected G1 in DMSO with concentrations of 50, 100, 150, and 200 mM. DMSO showed five main peaks at 4900, 4814, 4772, 4695, and 4618 cm^{-1}, which originated from the combinations of three out of the five fundamental vibrational modes, including C-S-C stretching, C-H bending, S=O stretching,

CH_3 rocking, and C-H stretching modes. Tables S1 and S2 in SI 2 summarize the peak assignments of the bands obtained for DMSO in the mid-IR region and those obtained for DMSO in the 5000-4500 cm^{-1} region, respectively.[33] With the increase in the G1 concentration, the mole fraction of DMSO decreases, and thus, the peaks whose second derivative intensities decreased as the glycine concentration increased were largely attributed to DMSO. From this perspective, the peaks at 4900, 4814, and 4772 cm^{-1} were assigned dominantly to DMSO. Spectral differences due to the species of the protecting groups existed in the 4700-4600 cm^{-1} region. In the case of the Fmoc group, the increase in the second derivative intensity at 4650 and 4600 cm^{-1} and the decrease in the second derivative intensity at 4695 and 4618 cm^{-1} of the second derivative intensities were observed, but in the case of the Boc group, on the other hand, such variations were not seen (Figure 5A and 5B). The two peaks at 4650 and 4600 cm^{-1} were assigned to the combination of C-H and C-C stretching modes of the benzene ring,[24] and the spectral differences that depend on the species of the protection groups were likely to be derived from the aromatic rings of the Fmoc group. Figure 5C examines the influence of the chain length of polyglycine. The intensity of the shoulder

at 4840 cm^{-1} and the one at approximately 4600 cm^{-1} seemed to systematically increase with increasing the chain length; these bands at 4840 cm^{-1} and approximately 4600 cm^{-1} can be assigned to the combination of amide A and amide II modes and that of amide A and amide III modes, respectively.[26, 27] Note that the spectral variations at 4600 cm^{-1} depended greatly on the types of protecting groups. This is considered to be due in part to the fact that the protective group with a cyclic structure had a large contribution to the band at 4600 cm^{-1} in addition to the amide bond. On the other hand, other bands except for the 4840 cm^{-1} band that originated from the combination of amide A and amide III modes did not seem to overlap at approximately 4840 cm^{-1}. Therefore, the band at 4840 cm^{-1} is more suitable than the other bands for use as a key band for the quantitative evaluation of the number of amide bonds.

Quantitative evaluation of the number of the amide bonds by PLSR

The number of amide bonds was quantitatively evaluated by PLSR using the second derivative spectral data obtained in the 4900-4500 cm^{-1} region for the polyglycine series, including Fmoc-G(n)-OH (n=1-3) and Fmoc-G(n)-6PG

(n=1-5) with four different concentrations: 50, 100, 150, and 200 mM. For example, Fmoc-G1-OH has one amide bond in the part connecting it to the Fmoc group, even though the side of the *N*-terminus was coupled with an oxygen atom, not a carbon atom similar to in -O-[C(=O)-NH]-C, and Fmoc-G2-OH has an additional amide bond in part of polyglycine (Figure 4A). The number of amide bonds in the 50 mM Fmoc-G1-OH solution was defined as 1, and the relative number of amide bonds in the 50 mM solutions could be assigned, as shown in Table 1. When the concentration of the solute was doubled or tripled, the relative number of amide bonds double or tripled, respectively. Furthermore, to investigate the differences between whether the coupling destination of the amide bond was an oxygen or a carbon atom, the number of amide bonds was evaluated separately for three cases; (i) *N*-terminus, (ii) polyglycine, and (iii) the total number of (i) and (ii). The definition of the amide bonds (i)-(iii) in Fmoc-G2-OH is demonstrated as an example in Figure 4A. Calibration models were built based on a leave-one-out cross validation method. The spectral data was left out for validation, and the calibration model was built using the remaining data. The validation data was applied to the model, and the prediction residuals were calculated. This

process was repeated until all the samples were excluded once. The accuracy of the models was assessed by the squared correlation coefficient (R^2) and root mean square error (RMSE).

PLSR was carried out for the three cases of the numbers of amide bonds, (i)-(iii), under the following wavenumber restrictions; 4900-4500 cm^{-1} (including both the bands due to the combination of the amide A and amide II/III modes) and 4900-4800 cm^{-1} (including only the band assigned to the combination of the amide A and amide II modes), and the results are demonstrated in Table 2. The calibration and validation accuracies were revealed to be better when focusing on the 4900-4800 cm^{-1} region than when it was not focused on. Since the peaks due to the Fmoc group had large contributions in the 4700-4600 cm^{-1} region due to the combination of C-H and C-C stretching modes in the benzene ring, the indeterminacy of the small peak intensities derived from the combination of the amide A and amide III modes at approximately 4618 cm^{-1} was likely to cause the quantitative accuracy to decrease.

Figures 6A and 6B show the loading plots of Factor 1 and Factor 2 of PLSR in the case of (i) under the restriction of the 4900-4800 cm^{-1} region,

where a band due to the combination of amide A and amide II modes appeared, respectively. Note that Factor 1, the component with the largest contribution, correlating with the number of amide bonds in all three cases ((i) 51.8%, (ii) 94.4%, and (iii) 95.2%), showed the same distributions as those shown in Figure 6A. Since Factor 1 had very high contributions in the cases of (ii) and (iii), it was likely to express the amide bond of polyglycine, and Factor 2 of (i) (contributing 36.2%) could be understood as a factor used to correct Factor 1 to show the amide bond in the *N*-terminus (Figure 6B).

To extract the element of the amide bond within the (i) *N*-terminus and (ii) polyglycine, the second derivative spectra of the 100 mM solution of Fmoc-G2-OH and the 50 mM solution of Fmoc-G3-OH were compared, for example. At first, the second derivative spectrum of pure DMSO was subtracted from that of the two organic solutions (Figure 6C), and the components that originated from each solute within the second derivative spectra were extracted (Figure 6D). The relative numbers of amide bonds at each concentration were (i, ii) = (2, 2) and (1, 2) for Fmoc-G2-OH and Fmoc-G3-OH, respectively. Therefore, the subtraction of the Fmoc-G3-OH spectra from those of Fmoc-G2-OH indicated that the component of the (i) *N*-

terminus amide bond could be extracted. Furthermore, the subtraction of this *N*-terminus amide bond component from Fmoc-G3-OH, in principle, showed that there were two amide bonds within (ii) polyglycine. Figures 6E and 6F depict the calculated second derivative spectra of the (ii) polyglycine and (i) *N*-terminus amide bonds in the 4900-4800 cm^{-1} region, respectively, as extracted by following the abovementioned procedures. The peak position of the amide band arising from the (i) *N*-terminus in the second derivative spectra (Figure 6F) shifted to a lower wavenumber than that from (ii) polyglycine (Figure 6E). Changing an atom attached to the amide bond in the *N*-terminus from carbon to oxygen may be a possible candidate of inducing the peak shift of the amide bond. In this way, the results of the PLSR of a series with Fmoc groups revealed that the elongation of the chain length of polyglycine was able to be estimated using the second derivative intensity due to the amide band in the 4900-4800 cm^{-1} region being obtained with a high accuracy, and the detailed peak features of the amide bands due to the (i) *N*-terminus and (ii) polyglycine amide bond were deeply understood by considering the meaning of the contributing factors in PLSR.

In a series of Boc groups, the relative number of amide bonds

classified as (iii) in the abovementioned definition was quantitatively evaluated. In this series, Boc-G(n)-OH (n=1-3) and Boc-G2-X-OH, where X was Ala, Ser(tBu), Glu(OtBu), Lys(Boc), Phe, and Tyr(tBu), were tried. The Xs had characteristic structures of amino acids. In particular, X=Lys(Boc) had an amide bond in the side chain, and X=Phe and Thy(tBu) had a ring structure. It was examined whether the number of amide bonds could be quantitatively evaluated, even though the amino acids exhibited structural differences. The evaluation accuracies of the numbers of amide bonds are exhibited in Table 3 separately for the two cases; whether an amino acid X was (a) without or (b) with a ring structure. In the case of (a) in the 4900-4500 cm^{-1} region, the contribution of the explained variance of Factor 1 was 94.4%, and two peaks were exhibited at 4842 and 4597 cm^{-1} due to combinations of amide A and amide II modes and combinations of amide A and amide III modes, respectively (Figure 7A). Both the calibration and validation accuracies were likely to be improved because the bands derived from the aromatic rings appearing in the 4700-4600 cm^{-1} region did not exist unlike in the Fmoc series. It seemed reasonable that the quantitative accuracies increased when two amide bands attributed to the combination of

the amide A and amide II/III modes were considered for carrying out PLSR.

In the case of (b), the contributions of the explained variances of Factor 1 and Factor 2 were 72.0 and 23.5%, respectively, summing up to 95.8%. Factor 1 included the peak arising from the amide band at 4842 cm^{-1}, and the peaks due to the aromatic ring components appeared at 4670 and 4574 cm^{-1}, with an artificial peak at 4611 cm^{-1} (Figure 7B), which merely appeared by subtracting the second derivative spectra from each other and was thus not a real peak. The loading plot of Factor 2, as the correcting component for Factor 1, contained an amide band at 4594 cm^{-1} in addition to the band at 4848 cm^{-1} (Figure 7B). Inclusion up to Factor 5 corresponding to a 99.6% explained variance gave the calibration and validation accuracies with R^2=0.99. Due to the wavenumber restriction, as the 4900-4800 cm^{-1} region only included peaks attributed to the combination of amide A and amide II modes, the quantitative accuracies decreased because of the reduction of the amount of useful information for estimating the number of amide bonds. Therefore, considering both amide bands at 4842 and 4594 cm^{-1} was clearly more important in the Boc group series than in the other series.

To verify the applicability of the calibration model between protection

groups, the dataset of the Boc group was applied as unknown samples to the calibration model built using the dataset of the Fmoc group. The validation results became R^2=0.83 and R^2=0.76 in the 4900-4500 and 4900-4800 cm^{-1} regions, respectively. Figure 8 depicts the plots of the calibration and validation results of the relative number of amide bonds using the data in the 4900-4800 cm^{-1} region. The estimated number of amide bonds was revealed to be systematically lower in the Boc group series than in the Fmoc group series. The reason for the underestimation was likely due to the peak position and peak form of the second derivative spectra showing the amide band in the 4900-4800 cm^{-1} region. The peak position of the amide band in the loading plots of Factor 1 for the Fmoc and Boc groups were 4864 and 4842 cm^{-1}, as shown in Figure 6A and Figure 7, respectively. Therefore, the calibration model is better to be built for each protecting group, or the modified model of one protecting group is applied to the other protecting group. These by NIR spectroscopy methods enable the quantitative evaluation of the number of amide bonds using the peak intensity variation of the amide bonds during the process of polypeptide synthesis.

Conclusion

A method monitoring the number of amide bonds increasing with the elongation of the chain length in polypeptide amino acids was investigated using NIR spectroscopy. The key absorption bands sensitively responding to the elongation of the chain length of polypeptide amino acids such as glycine, diglycine, triglycine, and tetraglycine were explored, and the quantitative accuracy of the number of amide bonds using the key bands was evaluated by PLSR. The results revealed that the combination of the amide A and II/III modes became key bands, giving a correlation coefficient higher than 0.99. The detailed peak features of the amide bands due to (i) *N*-terminus and (ii) polyglycine amide bonds were extensively studied by considering the meaning of the loadings of the contributed factors of PLSR, and the universal calibration model used for determining the number of amide bonds could be built by combining it with the modification depending on the species of protection groups. The present results not only have demonstrated the usefulness of NIR spectroscopy as a PAT tool for studying peptide synthesis in micro flow reactors but also have provided basic knowledge for analyzing amide bonds in the NIR spectra of protein, polyamino acids, polypeptides,

and polyamides. Our basic knowledge about the amide bands is extremely important and can be expanded to bioscience, agriculture, food engineering, and polymer engineering.

Acknowledgments

This work was supported by JST-Mirai Program Grant Number JPMJMI18G7, Japan

Supporting Information

Information on the polyglycine chemical reagents and supporting tables for the peak assignment of DMSO.

References

1. Castanho, M.; Nuno S. (Eds.) *Peptide drug discovery and development: translational research in academia and industry*, John Wiley & Sons, New Yolk, 2011.
2. Fuse, S.; Mifune, Y.; Takahashi, T. Efficient amide bond formation through a rapid and strong activation of carboxylic acids in a microflow reactor. *Angew. Chem. Int. Edi.*, **2014**, *53*, 851-855.
3. Fuse, S.; Otake, Y.; Nakamura, H. Peptide synthesis utilizing micro‐flow technology. *Chem. Asian J.*, **2018**, *13*, 3818-3832.
4. Ramesh, S.; Cherkupally, P.; Beatriz, G.; Govender, T.; Kruger, H. G.; Albericio, F. Microreactors for peptide synthesis: looking through the eyes of twenty first century!!!. *Amino acids*, **2014**, *46*, 2091-2104.
5. Myerson, A. S.; Krumme, M.; Nasr, M.; Thomas, H.; Braatz, R. D. Control systems engineering in continuous pharmaceutical manufacturing May 20–21, 2014 continuous manufacturing symposium. *J. Pharm. Sci.*, **2015**, *104*, 832-839.
6. Lee, S. L.; O'Connor, T. F.; Yang, X.; Cruz, C. N.; Chatterjee, S.; Madurawe, R. D.; Moore, C. M. V.; Yu, L. X.; Woodcock, J. Modernizing pharmaceutical manufacturing: from batch to continuous production. *J. Pharm. Innov.*, **2015**, *10*, 191-199.
7. Allison, G. et al. Regulatory and quality considerations for continuous manufacturing. May 20–21, 2014 continuous manufacturing symposium. *J. Pharm. Sci.*, **2015**, *104*, 803-812.
8. Pattabiraman, V. R.; Bode, J. W. Rethinking amide bond synthesis.

Nature, **2011**, *480*, 471-479.

9. Constable, D. J. et al. Key green chemistry research areas—a perspective from pharmaceutical manufacturers. *Green Chem.*, **2007**, *9*, 411-420.
10. Food and Drug Administration. *PAT—Framework for Innovative-Pharmaceutical Development, Manufacturing, and Quality Assurance. 2004*, 2004.
11. Bakeev, K. A. (Ed.). *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. John Wiley & Sons, New Yolk, 2010.
12. Harms, Z. D.; Shi, Z.; Kulkarni, R. A.; Myers, D. P. Characterization of near-infrared and Raman spectroscopy for in-line monitoring of a low-drug load formulation in a continuous manufacturing process. *Anal. Chem.*, **2019**, *91*, 8045-8053.
13. Esmonde-White, K. A.; Cuellar, M.; Uerpmann, C.; Lenain, B.; Lewis, I. R. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Anal. Bioanal. Chem.*, **2017**, *409*, 637-649.
14. Blanco, M.; Alcalá, M.; González, J. M.; Torras, E. A process analytical technology approach based on near infrared spectroscopy: tablet hardness, content uniformity, and dissolution test measurements of intact tablets. *J. Pharm. Sci.*, **2006**, *95*, 2137-2144.
15. Shah, R. B.; Tawakkul, M. A.; Khan, M. A. Process analytical technology: chemometric analysis of Raman and near infra-red spectroscopic data for predicting physical properties of extended release

matrix tablets. *J. Pharm. Sci.*, **2007**, *96*, 1356-1365.

16. Murayama, K.; Ishikawa, D.; Genkawa, T.; Ozaki, Y. An application for the quantitative analysis of pharmaceutical tablets using a rapid switching system between a near-infrared spectrometer and a portable near-infrared imaging system equipped with fiber optics. *Appl. Spectrosc.*, **2018**, *72*, 551-561.
17. Taday, P. F. Applications of terahertz spectroscopy to pharmaceutical sciences. *Philos. T. R. Soc. A*, **2004**. *362*, 351-364.
18. Siesler, H. W.; Ozaki, Y.; Kawata, S.; Heise, H. M. (Eds.). *Near-infrared spectroscopy*, Wiley-VCH, Weinheim, 2002.
19. Ozaki, Y.; McClure, W. F.; Christy, A. A. (Eds.). *Near-infrared spectroscopy in food science and technology*, John Wiley & Sons, New Yolk, 2006.
20. Ozaki, Y.; Huck, C. W.; Beć, K. B. Near infrared spectroscopy and its applications. In *Molecular and Laser Spectroscopy*, Gupta, V. P. (Ed.); Elsevier, Amsterdam, 2017; pp 11-38.
21. Ozaki, Y. Near-Infrared Spectroscopy – Its Versatility in Analytical chemistry. *Anal. Sci.* **2012**, *28*, 545–563.
22. Jue, T.; Masuda, K. (Eds.). Application of Near Infrared Spectroscopy in Biomedicine (Handbook of Modern Biophysics), Springer, Luxembourg, 2016.
23. Ishigaki, M.; Ozaki, Y. Near-infrared spectroscopy and imaging in protein research. In *Vibrational Spectroscopy in Protein Research*, Ozaki, Y. Baranska, M.; Lednev, I. K.; Wood, B. R. (Eds.), Elsevier, Amsterdam,

2020; 143-176.

24. Workman Jr, J.; Weyer, L. (Eds.). *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*, CRC Press, Boca Raton, 2012.
25. Robert, P.; Devaux, M. F.; Mouhous, N.; Dufour, E. Monitoring the secondary structure of proteins by near-infrared spectroscopy. *Appl. Spectrosc.*, **1999**, *53*, 226-232.
26. Holly, S.; Egyed, O.; Jalsovszky, G. Spectrochim. Acta A: *Mol. Spectrosc.*, **1992**, *48*, 101-109.
27. Yang, W. Y.; Larios, E.; Gruebele, M. J. *Am. Chem. Soc.*, **2003**, *125*, 16220-16227.
28. Ishigaki, M.; Yasui, Y.; Kajita, M.; Ozaki, Y. (2020). Assessment of Embryonic Bioactivity through Changes in the Water Structure Using Near-Infrared Spectroscopy and Imaging. *Anal. Chem.*, **2020**, *92*, 8133-8141.
29. Izutsu, K. I.; Fujimaki, Y.; Kuwabara, A.; Hiyama, Y.; Yomota, C.; Aoyagi, N. Near-infrared analysis of protein secondary structure in aqueous solutions and freeze-dried solids. *J. Pharm. Sci.*, **2006**, *95*, 781-789.
30. Miyazawa, M.; Sonoyama, M. Second derivative near infrared studies on the structural characterisation of proteins. *J. Near Infrared Spectros.*, **1998**, *6*, 253-257.
31. Hug, W.; Chalmers, J. M.; Griffith, P. R. (Eds.). *Handbook of vibrational spectroscopy*, John Wiley & Sons, New Yolk, 2002.
32. Damodaran, S. Amino acids, peptides and proteins. In *Fennema's food*

chemistry, 4th edition, Damodaran, S.; Parkin, K. L.; Fennema, O. R. (Eds.). CRC Press, Boca Raton, 2008; 217-329

33. Wallace, V. M.; Dhumal, N. R.; Zehentbauer, F. M.; Kim, H. J.; Kiefer, J. Revisiting the aqueous solutions of dimethyl sulfoxide by spectroscopy in the mid-and near-infrared: Experiments and Car–Parrinello simulations. *J. Phys. Chem. B*, **2015**, *119*, 14780-14789.

Table 1. The relative number of amide bonds in the 50 mM solutions. The number of amide bonds of Fmoc-G1-OH in the 50 mM solution was defined as 1.

	(i) Amide bond in *N*-terminus	(ii) Amide bond in polyglycine	(iii) Total number of amide bonds
Fmoc-G1-OH	1	0	1
Fmoc-G2-OH	1	1	2
Fmoc-G3-OH	1	2	3
Fmoc-G1-6PEG	1	0	1
Fmoc-G2-6PEG	1	1	2
Fmoc-G3-6PEG	1	2	3
Fmoc-G4-6PEG	1	3	4
Fmoc-G5-6PEG	1	4	5
Boc-G1-OH	1	0	1
Boc-G2-OH	1	1	2
Boc-G3-OH	1	2	3
Boc-G2-Ala-OH	1	2	3
Boc-G2-Ser(tBu)-OH	1	2	3
Boc-G2-Glu(OtBu)-OH	1	2	3
Boc-G2-Lys(Boc)-OH	1	3	4
Boc-G2-Phe-OH	1	2	3
Boc-G2-Phe-Tyr(tBu)-OH	1	2	3

Table 2. Results of PLSR for the Fmoc group. The number of amide bonds in (i) the *N*-terminus and (ii) polyglycine, and (iii) the total number of (i) and (ii).

	Spectral region (cm^{-1})	Calibration	Validation	Variables
(i)	4900-4500	R^2=0.980 RMSE=0.165	R^2=0.970 RMSE=0.211	5
	4900-4800	R^2=0.985 RMSE=0.144	R^2=0.975 RMSE=0.198	5
(ii)	4900-4500	R^2=0.985 RMSE=0.490	R^2=0.980 RMSE=0.577	3
	4900-4800	R^2=0.994 RMSE=0.310	R^2=0.990 RMSE=0.420	4
(iii)	4900-4500	R^2=0.985 RMSE=0.582	R^2=0.980 RMSE=0.676	2
	4900-4800	R^2=0.991 RMSE=0.450	R^2=0.988 RMSE=0.537	2

Table 3. Results of PLSR for the Boc group. The number of amide bonds (a) without a ring structure (Phe and Thy(tBu)) and (b) with it.

	Spectral region (cm^{-1})	Calibration	Validation	Variables
(a)	4900-4500	R^2=0.998 RMSE=0.175	R^2=0.993 RMSE=0.329	3
	4900-4800	R^2=0.994 RMSE=0.309	R^2=0.987 RMSE=0.465	4
(b)	4900-4500	R^2=0.996 RMSE=0.254	R^2=0.992 RMSE=0.361	5
	4900-4800	R^2=0.963 RMSE=0.748	R^2=0.948 RMSE=0.879	3

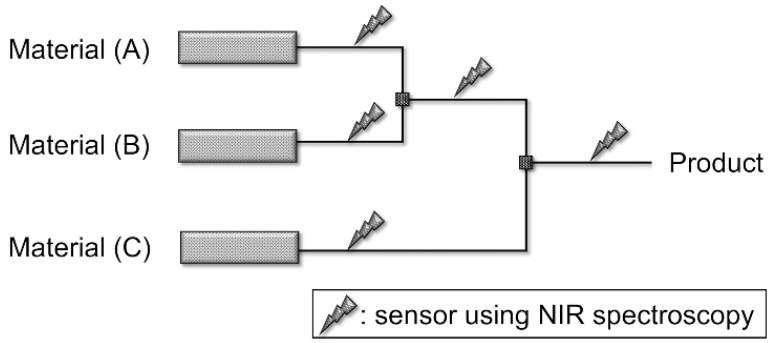

Figure 1: Schematic view of a microflow reactor.

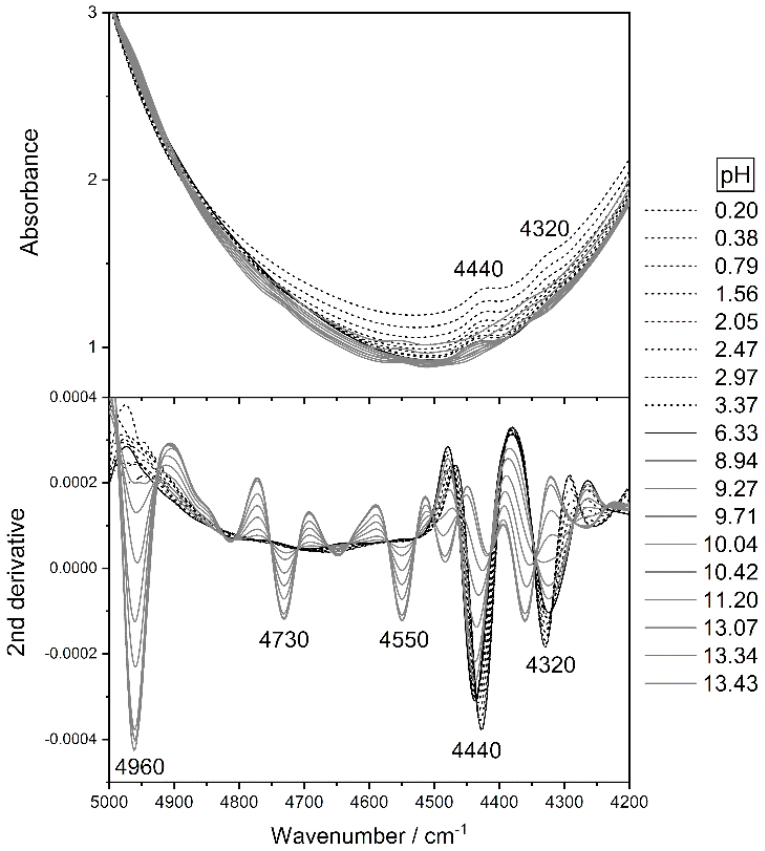

Figure 2: (A) NIR absorbance spectra in the 5000-4200 cm^{-1} region for the glycine aqueous solutions with a pH range of 0.20-13.43. (B) Second derivative spectra of (A).

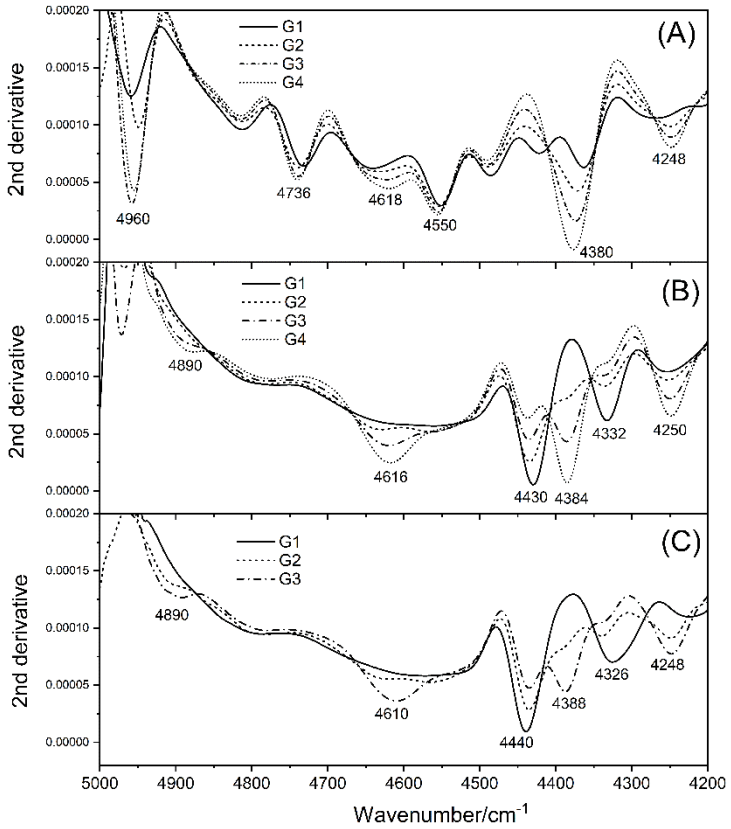

Figure 3: Second derivative spectra in the 5000-4200 cm^{-1} region for a series of polyglycine (G1, G2, G3, and G4) in (A) 0.5 mol/L NaOH, (B) 0.5 mol/L HCl, and (C) pure water.

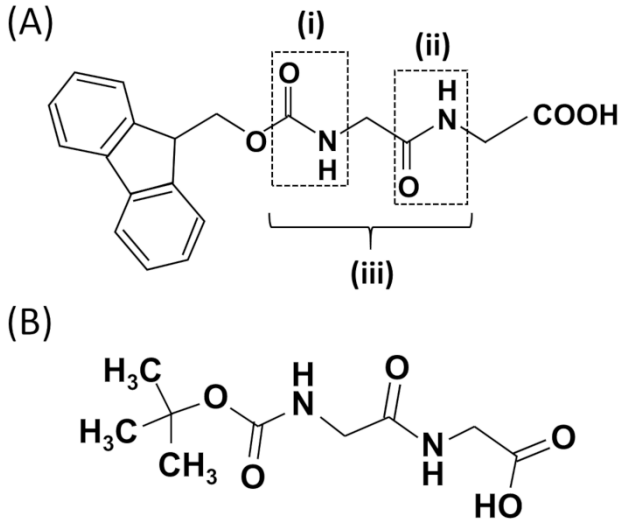

Figure 4: Chemical structures of (A) Fmoc-G2-OH and (B) Boc-G2-OH. The definition of the amide bond (i)-(iii) in Fmoc-G2-OH; (i) the *N*-terminus, and (ii) polyglycine, and (iii) the total number of (i) and (ii).

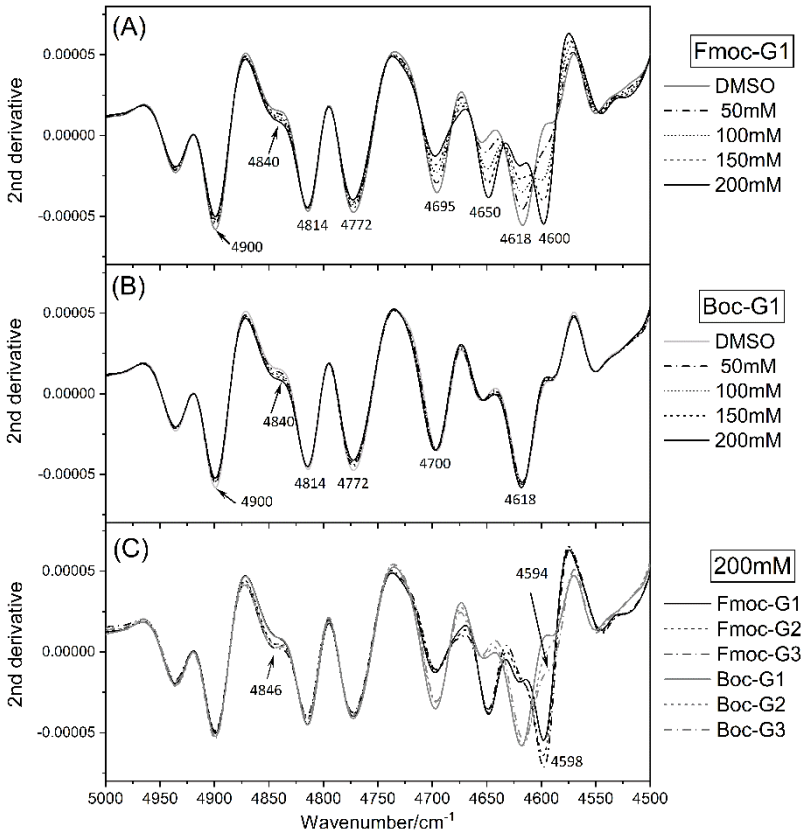

Figure 5: Second derivative spectra in the 5000-4500 cm^{-1} region of (A) Fmoc-protected and (B) Boc-Protected G1 in 25, 50, 100, 150, and 200 mM solutions. (C) The comparison of the second derivative spectra of the Fmoc- and Boc-protected G1-G3 in a 200 mM solution.

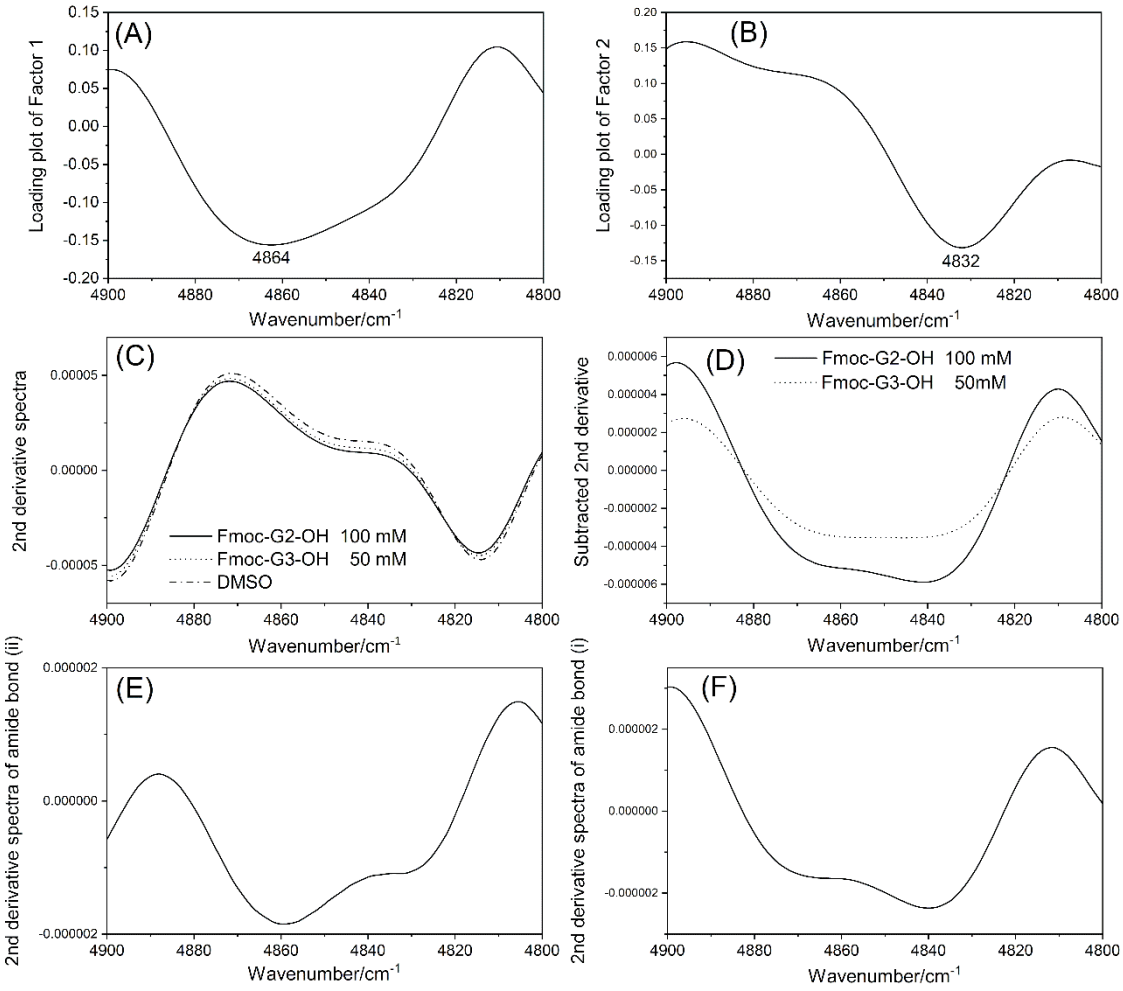

Figure 6: Loading plots of (A) Factor 1 and (B) Factor 2 of PLSR in the case of (i) under the restriction of the 4900-4800 cm^{-1} region. (C) Second derivative spectra in the 4900-4800 cm^{-1} region for Fmoc-G2-OH (100 mM), Fmoc-G3-OH (50 mM), and DMSO. (D) The second derivative spectra of Fmoc-G2-OH (100 mM) and Fmoc-G3-OH (50 mM) obtained by subtracting the second derivative spectrum of DMSO. The second derivative spectral components of the amide bonds within (E) polyglycine and the (F) *N*-terminus.

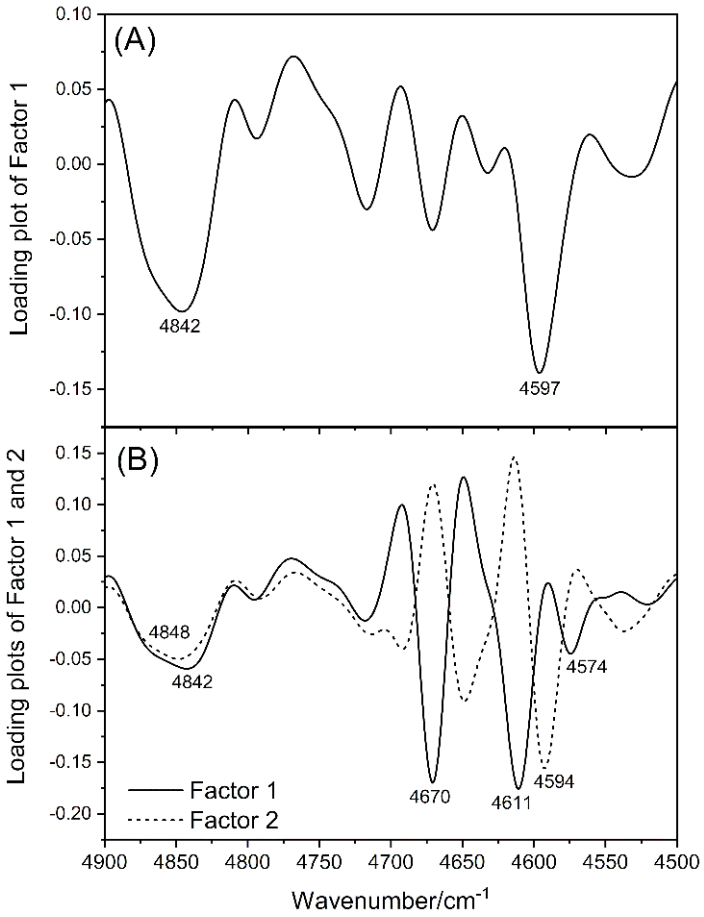

Figure 7: (A) A loading plot of Factor 1 in the case of (a) the 4900-4500 cm^{-1} region. (B) A loading plots of Factor 1 and 2 in the case of (b) the 4900-4500 cm^{-1} region.

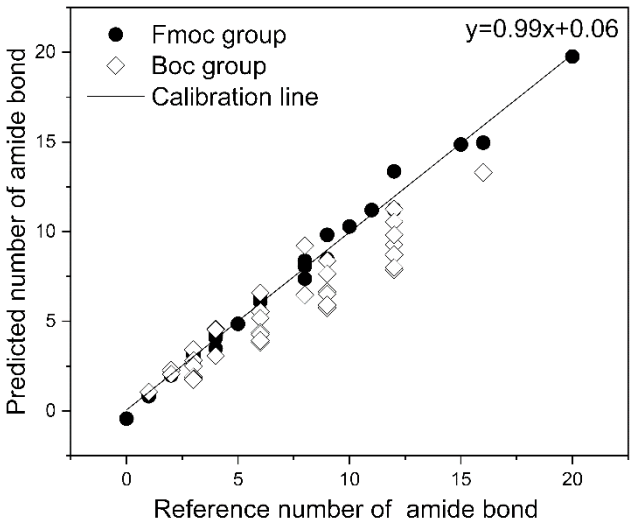

Figure 8: Plots of the calibration (Fmoc group) and validation (Boc group) results of the relative number of amide bonds obtained using the data in the 4900-4800 cm^{-1} region.

Supporting Information

A method of monitoring the number of amide bonds in peptides using near-infrared spectroscopy

Mika Ishigaki[1,2]*, Atsushi Ito[3], Risa Hara[3], Shun-ichi Miyazaki[3]*, Koudai Murayama[3], Keisuke Yoshikiyo[1], Tatsuyuki Yamamoto[1,2], Yukihiro Ozaki[4]

[1]*Institute of Agricultural and Life Sciences, Academic Assembly, Shimane University, 1060 Nishikawatsu, Matsue, Shimane, 690-8504, Japan*

[2]*Raman Project Center for Medical and Biological Applications, Shimane University, 1060 Nishikawatsu, Matsue, Shimane 690-8504, Japan*

[3]*Research and Development Department, Yokogawa Electric Corporation, 2-9-32 Nakacho, Musashino, Tokyo 180-8750, Japan*

[4]*School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan*

*Authors to whom correspondence should be sent.

*E-mail: ishigaki@life.shimane-u.ac.jp (M.I.)

shun-ichi.miyazaki@yokogawa.com (S.M)

Table of Contents

SI 1. Polyglycine chemical reagents

SI 2. Peak assignments of DMSO

SI 1. Polyglycine chemical reagents

Detailed information about the polyglycine chemical reagents is presented.

(i) Fmoc (9-fluorenylmethyloxycarbonyl group)-Protected G(n)-OH

- Fmoc-G1-OH: 29022-11-5, Watanabe Chemical Ind. Ltd., Japan
- Fmoc-G2-OH: 35665-38-4, Watanabe Chemical Ind. Ltd., Japan
- Fmoc-G3-OH: 170941-79-4, Watanabe Chemical Ind. Ltd., Japan

(ii) Fmoc-Protected G(n) with 6-Polyethylene glycol (6-PEG) (n=1-5)

- Fmoc-G1-6PEG-OH: M_W=632.5, >90%, HiPep Laboratories Kyoto, Japan
- Fmoc-G3-6PEG: M_W=746.80, >90%, HiPep Laboratories Kyoto, Japan
- Fmoc-G4-6PEG-OH: M_W=803.6, >90%, HiPep Laboratories Kyoto, Japan
- Fmoc-G5-6PEG-OH: M_W=860.72, >90%, HiPep Laboratories Kyoto, Japan

(iii) Boc (N-tert-Butoxycarbonyl)-Protected G(n)-OH

- Boc-G1-OH: 4530-20-5, Watanabe Chemical Ind. Ltd., Japan
- Boc-G2-OH: 31972-52-8, Watanabe Chemical Ind. Ltd., Japan
- Boc-G3-OH: 28320-73-2, Watanabe Chemical Ind. Ltd., Japan

(iv) Boc-Protected diglycine with X

- Boc-GlyGlyAla-OH: M_w=303.31, >90%, HiPep Laboratories Kyoto, Japan
- Boc-GlyGlySer(tBu)-OH: M_w=375.29, >80%, HiPep Laboratories Kyoto, Japan
- Boc-GlyGlyGlu(OtBu)-OH: M_w=417.32, >80%, HiPep Laboratories Kyoto, Japan
- Boc-GlyGlyLys(Boc)-OH: M_w=460.39, >80%, HiPep Laboratories Kyoto, Japan
- Boc-GlyGlyPhe-OH: M_w=379.41, >90%, HiPep Laboratories Kyoto, Japan
- Boc-GlyGlyTyr(tBu)-OH: M_w=451.38, >80%, HiPep Laboratories Kyoto, Japan

SI 2: Peak assignments of DMSO

Table S1 and S2 summarize the peak assignments of the bands due to the fundamental modes of DMSO in the mid-IR region and those obtained for DMSO in the 5000-4500 cm^{-1} region, respectively.

Table S1. Peak assignments of an IR spectrum of DMSO.[33]

	Peak Position (cm^{-1})	Peak Assignment
i	667, 697	C-S-C stretching
ii	952, 1310, 1407	C-H bending
iii	1041, 1052	S=O stretching
iv	1019, 1031	rocking vibration of CH_3
v	2912, 2996	C-H stretching

Table S2. Possible peak assignments of a NIR spectrum in the 4900-4500 cm^{-1} region of DMSO.

Peak Position (cm^{-1})	Peak Assignment
4900	Comb. of (v, iv, ii), (v, iii, ii), (v, ii, i), (v, 2×ii)
4814,	Comb. of (2×ii,v)
4772	Comb. of (v, iv, i), (v, iii, i)
4695	Comb. of (v, iv, i), (v, iii, i)
4618	Comb. of (v, iv, i), (v, iii, i), (v, ii, i)

Reference

33. Wallace, V. M.; Dhumal, N. R.; Zehentbauer, F. M.; Kim, H. J.; Kiefer, J. Revisiting the aqueous solutions of dimethyl sulfoxide by spectroscopy in the mid-and near-infrared: Experiments and Car–Parrinello simulations. *J. Phys. Chem. B*, **2015**, *119*, 14780-14789.